

Document AP1
Appl. No. 09/891,332

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/11, C12P 19/34, C12Q 1/68		A1	(11) International Publication Number: WO 95/13369
			(43) International Publication Date: 18 May 1995 (18.05.95)
(21) International Application Number: PCT/US94/13041		(81) Designated States: AU, CA, FI, JP, NO, SI, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 14 November 1994 (14.11.94)			
(30) Priority Data: 152,482 12 November 1993 (12.11.93) US		Published <i>With international search report.</i>	
(71) Applicant: THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).			
(72) Inventors: ERLANDER, Mark, G.; 1352 Via Terrassa, Encinitas, CA 92024 (US). SUTCLIFFE, Gregor, J.; 2253 Via Tiempo, Cardiff, CA 92007 (US).			
(74) Agent: FITTING, Thomas; The Scripps Research Institute, Patent Counsel, TPC-8, 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).			
(54) Title: METHOD FOR SIMULTANEOUS IDENTIFICATION OF DIFFERENTIALLY EXPRESSED mRNAs AND MEASUREMENT OF RELATIVE CONCENTRATIONS			
(57) Abstract			
<p>An improved method for the simultaneous sequence-specific identification of mRNAs in an mRNA population allows the visualization of nearly every mRNA expressed by a tissue as a distinct band on a gel whose intensity corresponds roughly to the concentration of the mRNA. In general, the method comprises the formation of cDNA using anchor primers to fix a 3'-endpoint, producing cloned inserts from the cDNA in a vector containing a bacteriophage-specific promoter for subsequent RNA synthesis, generating linearized fragments of the cloned inserts, preparing cRNA, transcribing cDNA from the cRNA using a set of primers, and performing PCR using a 3'-primer whose sequence is derived from the vector and a set of 5'-primers that is derived from the primers used for transcription of cDNA from cRNA. The method can identify changes in expression of mRNA associated with the administration of drugs or with physiological or pathological conditions.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China			TD	Chad

DK	Denmark	SL	Singapore	TM	Kazakhstan
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Malta	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

METHOD FOR SIMULTANEOUS IDENTIFICATION
OF DIFFERENTIALLY EXPRESSED mRNAs AND MEASUREMENT
OF RELATIVE CONCENTRATIONS

BACKGROUND OF THE INVENTION

This invention is directed to methods for simultaneous identification of differentially expressed mRNAs, as well as measurements of their relative concentrations.

5

An ultimate goal of biochemical research ought to be a complete characterization of the protein molecules that make up an organism. This would include their identification, sequence determination, 10 demonstration of their anatomical sites of expression, elucidation of their biochemical activities, and understanding of how these activities determine organismic physiology. For medical applications, the description should also include information about how the 15 concentration of each protein changes in response to pharmaceutical or toxic agents.

Let us consider the scope of the problem: How many genes are there? The issue of how many genes are 20 expressed in a mammal is still unsettled after at least two decades of study. There are few direct studies that address patterns of gene expression in different tissues. Mutational load studies (J.O. Bishop, "The Gene Numbers Game," Cell 2:81-86 (1974); T. Ohta & M. Kimura, 25 "Functional Organization of Genetic Material as a Product of Molecular Evolution," Nature 223:118-119 (1971)) have suggested that there are between 3×10^4 and 10^5 essential genes.

30 Before cDNA cloning techniques, information on

mixed populations of RNA molecules with different

- 2 -

specificities in abundances. To an unexpected extent, early analog complexity studies were distorted by hidden complications of the fact that the molecules in each tissue that make up most of its mRNA mass comprise only a 5 small fraction of its total complexity. Later, cDNA cloning allowed digital measurements (i.e., sequence-specific measurements on individual species) to be made; hence, more recent concepts about mRNA expression are based upon actual observations of individual RNA species.

10

Brain, liver, and kidney are the mammalian tissues that have been most extensively studied by analog RNA complexity measurements. The lowest estimates of complexity are those of Hastie and Bishop (N.D. Hastie & 15 J. B. Bishop, "The Expression of Three Abundance Classes of Messenger RNA in Mouse Tissues," Cell 9:761-774 (1976)), who suggested that 26×10^6 nucleotides of the 3×10^9 base pair rodent genome were expressed in brain, 23×10^6 in liver, and 22×10^6 in kidney, with nearly 20 complete overlap in RNA sets. This indicates a very minimal number of tissue-specific mRNAs. However, experience has shown that these values must clearly be underestimates, because many mRNA molecules, which were probably of abundances below the detection limits of this 25 early study, have been shown to be expressed in brain but detectable in neither liver nor kidney. Many other researchers (J.A. Bantle & W.E. Hahn, "Complexity and Characterization of Polyadenylated RNA in the Mouse Brain," Cell 8:139-150 (1976); D.M. Chikaraishi, 30 "Complexity of Cytoplasmic Polyadenylated and Non-Adenylated Rat Brain Ribonucleic Acids," Biochemistry 18:3249-3256 (1979)) have measured analog complexities of between $100-200 \times 10^6$ nucleotides in brain, and 2-to-3-fold

values have been supported by digital cloning studies.

-3-

(R.J. Milner & J.G. Sutcliffe, "Gene Expression in Rat Brain," Nucl. Acids Res. 11:5497-5520 (1983)).

Analog measurements on bulk mRNA suggested that
5 the average mRNA length was between 1400-1900 nucleotides. In a systematic digital analysis of brain mRNA length using 200 randomly selected brain cDNAs to measure RNA size by northern blotting (Milner & Sutcliffe, supra), it was found that, when the mRNA size
10 data were weighted for RNA prevalence, the average length was 1790 nucleotides, the same as that determined by analog measurements. However, the mRNAs that made up most of the brain mRNA complexity had an average length of 5000 nucleotides. Not only were the rarer brain RNAs
15 longer, but they tended to be brain specific, while the more prevalent brain mRNAs were more ubiquitously expressed and were much shorter on average.

These concepts about mRNA lengths have been
20 corroborated more recently from the length of brain mRNA whose sequences have been determined (J.G. Sutcliffe, "mRNA in the Mammalian Central Nervous System," Annu. Rev. Neurosci. 11:157-198 (1988)). Thus, the $1-2 \times 10^8$ nucleotide complexity and 5000-nucleotide average mRNA
25 length calculates to an estimated 30,000 mRNAs expressed in the brain, of which about 2/3 are not detected in liver or kidney. Brain apparently accounts for a considerable portion of the tissue-specific genes of mammals. Most brain mRNAs are expressed at low
30 concentration. There are no total-mammal mRNA complexity measurements, nor is it yet known whether 5000 nucleotides is a good mRNA-length estimate for non-neuronal tissues. A reasonable estimate of total gene number

- 4 -

What is most needed to advance by a chemical understanding of physiological function is a menu of protein sequences encoded by the genome plus the cell types in which each is expressed. At present, protein 5 sequences can be reliably deduced only from cDNAs, not from genes, because of the presence of the intervening sequences (introns) in the genomic sequences. Even the complete nucleotide sequence of a mammalian genome will not substitute for characterization of its expressed 10 sequences. Therefore, a systematic strategy for collecting transcribed sequences and demonstrating their sites of expression is needed. Such a strategy would be of particular use in determining sequences expressed differentially within the brain. It is necessarily an 15 eventual goal of such a study to achieve closure; that is, to identify all mRNAs. Closure can be difficult to obtain due to the differing prevalence of various mRNAs and the large number of distinct mRNAs expressed by many distinct tissues. The effort to obtain it allows one to 20 obtain a progressively more reliable description of the dimensions of gene space.

Studies carried out in the laboratory of Craig Venter (M.D. Adams et al., "Complementary DNA Sequencing: 25 Expressed Sequence Tags and Human Genome Project," Science 252:1651-1656 (1991); M.D. Adams et al., "Sequence Identification of 2,375 Human Brain Genes," Nature 355:632-634 (1992)) have resulted in the isolation of randomly chosen cDNA clones of human brain mRNAs, the 30 determination of short single-pass sequences of their 3'-ends, about 300 base pairs, and a compilation of some 2500 of these as a database of "expressed sequence tags." This database, while useful, fails to provide any

"overall pattern of expression within regions of brain and

-5-

other tissues and in response to various paradigms, such as various physiological or pathological states or the effects of drug treatment, rather than simply their expression in a single tissue.

5

Other work has focused on the use of the polymerase chain reaction (PCR) to establish a database. Williams et al. (J.G.K. Williams et al., "DNA Polymorphisms Amplified by Arbitrary Primers Are Useful 10 as Genetic Markers," Nucl. Acids Res. 18:6531-6535 (1990)) and Welsh & McClelland (J. Welsh & McClelland, "Genomic Fingerprinting Using Arbitrarily Primed PCR and a Matrix of Pairwise Combinations of Primers," Nucl. Acids Res. 18:7213-7218 (1990)) showed that single 10-mer 15 primers of arbitrarily chosen sequences, i.e., any 10-mer primer off the shelf, when used for PCR with complex DNA templates such as human, plant, yeast, or bacterial genomic DNA, gave rise to an array of PCR products. The priming events were demonstrated to involve incomplete 20 complementarity between the primer and the template DNA. Presumably, partially mismatched primer-binding sites are randomly distributed through the genome. Occasionally, two of these sites in opposing orientation were located closely enough together to give rise to a PCR product 25 band. There were on average 8-10 products, which varied in size from about 0.4 to about 4 kb and had different mobilities for each primer. The array of PCR products exhibited differences among individuals of the same species. These authors proposed that the single 30 arbitrary primers could be used to produce restriction fragment length polymorphism (RFLP)-like information for genetic studies. Others have applied this technology (S.R. Woodward et al., "Random Sequence Oligonucleotide

SEARCHED _____ SERIALIZED _____ INDEXED _____ FILED _____

J.H. Nadeau et al., "Multiplex Markers for Mouse Genome

- 6 -

Analysis: PCR Amplification Based on Single Primers of Arbitrary Nucleotide Sequence," Mamm. Genome 3:55-64 (1992)).

5 Two groups (J. Welsh et al., "Arbitrarily Primed PCR Fingerprinting of RNA," Nucl. Acids Res. 20:4965-4970 (1992); P. Liang & A.B. Pardee, "Differential Display of Eukaryotic Messenger RNA by Means of the Polymerase Chain Reaction," Science 257:967-971 (1992)) adapted the method to compare mRNA populations. In the study of Liang and Pardee, this method, called mRNA differential display, was used to compare the population of mRNAs expressed by two related cell types, normal and tumorigenic mouse A31 cells. For 10 each experiment, they used one arbitrary 10-mer as the 5'-primer and an oligonucleotide complementary to a subset of poly A tails as a 3' anchor primer, performing PCR amplification in the presence of ³⁵S-dNTPs on cDNAs prepared from the two cell types. The products were 15 resolved on sequencing gels and 50-100 bands ranging from 100-500 nucleotides were observed. The bands presumably resulted from amplification of cDNAs corresponding to the 3'-ends of mRNAs that contain the complement of the 3' anchor primer and a partially mismatched 5' primer site, 20 as had been observed on genomic DNA templates. For each primer pair, the pattern of bands amplified from the two cDNAs was similar, with the intensities of about 80% of the bands being indistinguishable. Some of the bands 25 were more intense in one or the other of the PCR samples; some were detected in only one of the two samples.

Further studies (P. Liang et al., "Distribution and Cloning of Eukaryotic mRNAs by Means of Differential

works with low concentrations of input RNA although it

- 7 -

is not quantitative for rarer species), and the specificity resides primarily in the last nucleotide of the 3' anchor primer. At least a third of identified differentially detected PCR products correspond to 5 differentially expressed RNAs, with a false positive rate of at least 25%.

If all of the 50,000 to 100,000 mRNAs of the mammal were accessible to this arbitrary-primer PCR 10 approach, then about 80-95 5' arbitrary primers and 12 3' anchor primers would be required in about 1000 PCR panels and gels to give a likelihood, calculated by the Poisson distribution, that about two-thirds of these mRNAs would be identified.

15

It is unlikely that all mRNAs are amenable to detection by this method for the following reasons. For an mRNA to surface in such a survey, it must be prevalent enough to produce a signal on the autoradiograph and 20 contain a sequence in its 3' 500 nucleotides capable of serving as a site for mismatched primer binding and priming. The more prevalent an individual mRNA species, the more likely it would be to generate a product. Thus, prevalent species may give bands with many different 25 arbitrary primers. Because this latter property would contain an unpredictable element of chance based on selection of the arbitrary primers, it would be difficult to approach closure by the arbitrary primer method. Also, for the information to be portable from one 30 laboratory to another and reliable, the mismatched priming must be highly reproducible under different laboratory conditions using different PCR machines, with the resulting slight variation in reaction conditions. As

obtained by the Liang & Pardee differential display method.

There is therefore a need for an improved
5 method of differential display of mRNA species that reduces the uncertain aspect of 5'-end generation and allows data to be absolutely reproducible in different settings. Preferably, such a method does not depend on potentially irreproducible mismatched priming.
10 Preferably, such a method reduces the number of PCR panels and gels required for a complete survey and allows double-strand sequence data to be rapidly accumulated. Preferably, such an improved method also reduces, if not eliminates, the number of concurrent signals obtained
15 from the same species of mRNA.

SUMMARY

We have developed an improved method for the simultaneous sequence-specific identification of mRNAs in
20 a mRNA population. In general, this method comprises:

(1) preparing double-stranded cDNAs from a mRNA population using a mixture of 12 anchor primers, the anchor primers each including: (i) a tract of from 7 to 40 T residues; (ii) a site for cleavage by a restriction
25 endonuclease that recognizes more than six bases, the site for cleavage being located to the 5'-side of the tract of T residues; (iii) a stuffer segment of from 4 to 40 nucleotides, the stuffer segment being located to the 5'-side of the site for cleavage by the restriction endonuclease; and (iv) phasing residues -V-N located at the 3' end of each of the anchor primers, wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from
30

- 9 -

(2) producing cloned inserts from a suitable host cell that has been transformed by a vector, the vector having the cDNA sample that has been cleaved with a first restriction endonuclease and a second restriction 5 endonuclease inserted therein, the cleaved cDNA sample being inserted in the vector in an orientation that is antisense with respect to a bacteriophage-specific promoter within the vector, the first restriction endonuclease recognizing a four-nucleotide sequence and 10 the second restriction endonuclease cleaving at a single site within each member of the mixture of anchor primers;

(3) generating linearized fragments of the cloned inserts by digestion with at least one restriction endonuclease that is different from the first and second 15 restriction endonucleases;

(4) generating a cRNA preparation of antisense cRNA transcripts by incubation of the linearized fragments with a bacteriophage-specific RNA polymerase capable of initiating transcription from the 20 bacteriophage-specific promoter;

(5) dividing the cRNA preparation into sixteen subpools and transcribing first-strand cDNA from each subpool, using a thermostable reverse transcriptase and one of sixteen primers whose 3'-terminus is -N-N, wherein 25 N is one of the four deoxyribonucleotides A, C, G, or T, the primer being at least 15 nucleotides in length, corresponding in sequence to the 3'-end of the bacteriophage-specific promoter, and extending across into at least the first two nucleotides of the cRNA, the 30 mixture including all possibilities for the 3'-terminal two nucleotides;

(6) using the product of transcription in each of the sixteen subpools as a template for a polymerase

of insertion of the cDNA sample in the vector and a 3

-10-

primer selected from the group consisting of: (i) the primer from which first-strand cDNA was made for that subpool; (ii) the primer from which the first-strand cDNA was made for that subpool extended at its 3'-terminus by 5 an additional residue -N, where N can be any of A, C, G, or T; and (iii) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can be any of A, C, G, or T, to produce polymerase chain 10 reaction amplified fragments; and

(7) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands representing the 3'-ends of mRNAs present in the sample.

15 Typically, the anchor primers each have 18 T residues in the tract of T residues, and the stuffer segment of the anchor primers is 14 residues in length. A suitable sequence for the stuffer segment is A-A-C-T-G-G-A-A-G-A-A-T-T-C (SEQ ID NO: 1).

20 Typically, the site for cleavage by a restriction endonuclease that recognizes more than six bases is the NotI cleavage site. In this case, suitable anchor primers have the sequence A-A-C-T-G-G-A-A-G-A-A-T-25 T-C-G-C-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2).

30 Typically, the bacteriophage-specific promoter is selected from the group consisting of T3 promoter and T7 promoter. Most typically, it is the T3 promoter.

Typically, the sixteen primers for priming of transcription of cDNA from cRNA have the sequence A-G-G-

-11-

The vector can be the plasmid pBC SK⁺ cleaved with ClaI and NotI, in which case the 3'-primer in step (6) can be G-A-A-C-A-A-A-G-C-T-G-G-A-G-C-T-C-C-A-C-C-G-C (SEQ ID NO: 4).

5

The first restriction endonuclease recognizing a four-nucleotide sequence is typically MspI; alternatively, it can be TagI or HinPII. The restriction endonuclease cleaving at a single site in each of the mixture of anchor primers is typically NotI.

Typically, the mRNA population has been enriched for polyadenylated mRNA species.

15 A typical host cell is a strain of Escherichia coli.

The step of generating linearized fragments of the cloned inserts typically comprises:

20 (a) dividing the plasmid containing the insert into two fractions, a first fraction cleaved with the restriction endonuclease XbaI and a second fraction cleaved with the restriction endonuclease SalI;

(b) recombining the first and second
25 fractions after cleavage;

(c) dividing the recombined fractions into thirds and cleaving the first third with the restriction endonuclease HindIII, the second third with the restriction endonuclease BamHI, and the third third with the restriction endonuclease EcoRI; and

(d) recombining the thirds after digestion in order to produce a population of linearized fragments of which about one-sixth of the population corresponds to

-12-

Typically, the step of resolving the polymerase chain reaction amplified fragments by electrophoresis comprises electrophoresis of the fragments on at least two gels.

5

The method can further comprise determining the sequence of the 3'-end of at least one of the mRNAs, such as by:

- (1) eluting at least one cDNA corresponding to 10 a mRNA from an electropherogram in which bands representing the 3'-ends of mRNAs present in the sample are displayed;
- (2) amplifying the eluted cDNA in a polymerase chain reaction;
- 15 (3) cloning the amplified cDNA into a plasmid;
- (4) producing DNA corresponding to the cloned DNA from the plasmid; and
- (5) sequencing the cloned cDNA.

20

Another aspect of the invention is a method of simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs, the antisense cRNAs that are members of the antisense cRNA 25 pool being terminated at their 5'-end with a primer sequence corresponding to a bacteriophage-specific vector and at their 3'-end with a sequence corresponding in sequence to a sequence of the vector. The method comprises;

30

(1) dividing the members of the antisense cRNA pool into sixteen subpools and transcribing first-strand cDNA from each subpool, using a thermostable reverse transcriptase and one of sixteen primers whose 3'-

least 11 nucleotides in length, corresponding in sequence

-13-

to the 3'-end of the bacteriophage-specific promoter, and extending across into at least the first two nucleotides of the cRNA, the mixture including all possibilities for the 3'-terminal two nucleotides;

5 (2) using the product of transcription in each of the sixteen subpools as a template for a polymerase chain reaction with a 3'-primer that corresponds in sequence to a sequence vector adjoining the site of insertion of the cDNA sample in the vector and a 5'-
10 primer selected from the group consisting of: (i) the primer from which first-strand cDNA was made for that subpool; (ii) the primer from which the first-strand cDNA was made for that subpool extended at its 3'-terminus by an additional residue -N, where N can be any of A, C, G,
15 or T; and (iii) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can be any of A, C, G, or T, to produce polymerase chain reaction amplified fragments; and

20 (3) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands representing the 3'-ends of mRNAs present in the sample.

Yet another aspect of the present invention is
25 a method for detecting a change in the pattern of mRNA expression in a tissue associated with a physiological or pathological change. This method comprises the steps of:

30 (1) obtaining a first sample of a tissue that is not subject to the physiological or pathological change;

 (2) determining the pattern of mRNA expression in the first sample of the tissue by performing steps (1)-(3) of the method described above for simultaneous

ends of a population of mRNAs to generate a first display

-14-

of bands representing the 3'-ends of mRNAs present in the first sample;

(3) obtaining a second sample of the tissue that has been subject to the physiological or
5 pathological change;

(4) determining the pattern of mRNA expression in the second sample of the tissue by performing steps (1)-(3) of the method described above for simultaneous sequence-specific identification of mRNAs corresponding
10 to members of an antisense cRNA pool to generate a second display of bands representing the 3'-ends of mRNAs present in the second sample; and

(5) comparing the first and second displays to determine the effect of the physiological or pathological
15 change on the pattern of mRNA expression in the tissue.

The comparison is typically made in adjacent lanes.

20 The tissue can be derived from the central nervous system or from particular structures within the central nervous system. The tissue can alternatively be derived from another organ or organ system.

25 Another aspect of the present invention is a method of screening for a side effect of a drug. The method can comprise the steps of:

(1) obtaining a first sample of tissue from an organism treated with a compound of known physiological
30 function;

(2) determining the pattern of mRNA expression in the first sample of the tissue by performing steps (1)-(3) of the method described above for simultaneous

-15-

display of bands representing the 3'-ends of mRNAs present in the first sample;

5 (3) obtaining a second sample of tissue from an organism treated with a drug to be screened for a side effect;

10 (4) determining the pattern of mRNA expression in the second sample of the tissue by performing steps (1)-(3) of the method described above for simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool to generate a second display of bands representing the 3'-ends of mRNAs present in the second sample; and

15 (5) comparing the first and second displays in order to detect the presence of mRNA species whose expression is not affected by the known compound but is affected by the drug to be screened, thereby indicating a difference in action of the drug to be screened and the known compound and thus a side effect.

20 The drug to be screened can be a drug affecting the central nervous system, such as an antidepressant, a neuroleptic, a tranquilizer, an anticonvulsant, a monoamine oxidase inhibitor, or a stimulant.

25 Alternatively, the drug can be another class of drug such as an anti-parkinsonism agent, a skeletal muscle relaxant, an analgesic, a local anesthetic, a cholinergic, an antispasmodic, a steroid, or a non-steroidal anti-inflammatory drug.

30 Another aspect of the present invention is panels of primers and degenerate mixtures of primers suitable for the practice of the present invention. These include:

NO: 3), wherein N is one of the four deoxyribonucleotides A, C, G, or T;

(2) a panel of primers comprising 64 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO; 5), wherein N is one of the four deoxyribonucleotides A, C, G, or T;

(3) a panel of primers comprising 256 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO: 6), wherein N is one of the four deoxyribonucleotides A, C, G, or T; and

(4) a panel of primers comprising 12 primers of the sequences A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T; and

(5) a degenerate mixture of primers comprising a mixture of 12 primers of the sequences A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T, each of the 12 primers being present in about an equimolar quantity.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present invention will become better understood with reference to the following description, appended claims, and accompanying drawings where:

Figure 1 is a diagrammatic depiction of the method of the present invention showing the various

-17-

Figure 2 is an autoradiogram of a gel showing the result of performing the method of the present invention using several 5'-primers in the PCR step corresponding to known sequences of brain mRNAs and using
5 liver and brain mRNA as starting material.

DESCRIPTION

We have developed a method for simultaneous sequence-specific identification and display of mRNAs in
10 a mRNA population.

As discussed below, this method has a number of applications in drug screening, the study of physiological and pathological conditions, and genomic
15 mapping. These applications will be discussed below.

I. SIMULTANEOUS SEQUENCE-SPECIFIC IDENTIFICATION OF mRNAs

A method according to the present invention,
20 based on the polymerase chain reaction (PCR) technique, provides means for visualization of nearly every mRNA expressed by a tissue as a distinct band on a gel whose intensity corresponds roughly to the concentration of the mRNA. The method is based on the observation that
25 virtually all mRNAs conclude with a 3'-poly (A) tail but does not rely on the specificity of primer binding to the tail.

In general, the method comprises:
30 (1) preparing double-stranded cDNAs from a mRNA population using a mixture of 12 anchor primers, the anchor primers each including: (i) a tract of from 7 to 40 T residues; (ii) a site for cleavage by a restriction

tract of T residues; (iii) a stuffer segment of from 4 to

40 nucleotides, the stuffer segment being located to the 5'-side of the site for cleavage by the restriction endonuclease; and (iv) phasing residues -V-N located at the 3' end of each of the anchor primers, wherein V is a 5 deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T, the mixture including anchor primers containing all possibilities for V and N;

10 (2) producing cloned inserts from a suitable host cell that has been transformed by a vector, the vector having the cDNA sample that has been cleaved with a first restriction endonuclease and a second restriction endonuclease inserted therein, the cleaved cDNA sample 15 being inserted in the vector in an orientation that is antisense with respect to a bacteriophage-specific promoter within the vector, the first restriction endonuclease recognizing a four-nucleotide sequence and the second restriction endonuclease cleaving at a single 20 site within each member of the mixture of anchor primers;

(3) generating linearized fragments of the cloned inserts by digestion with at least one restriction endonuclease that is different from the first and second restriction endonucleases;

25 (4) generating a cRNA preparation of antisense cRNA transcripts by incubation of the linearized fragments with a bacteriophage-specific RNA polymerase capable of initiating transcription from the bacteriophage-specific promoter;

30 (5) dividing the cRNA preparation into sixteen subpools and transcribing first-strand cDNA from each subpool, using a thermostable reverse transcriptase and one of sixteen primers whose 3'-terminus is -N-N, wherein

corresponding in sequence to the 3'-end of the

-19-

bacteriophage-specific promoter, and extending across into at least the first two nucleotides of the cRNA, the mixture including all possibilities for the 3'-terminal two nucleotides;

5 (6) using the product of transcription in each of the sixteen subpools as a template for a polymerase chain reaction with a 3'-primer that corresponds in sequence to a sequence in the vector adjoining the site of insertion of the cDNA sample in the vector and a 5'-
10 primer selected from the group consisting of: (i) the primer from which first-strand cDNA was made for that subpool; (ii) the primer from which the first-strand cDNA was made for that subpool extended at its 3'-terminus by an additional residue -N, where N can be any of A, C, G,
15 or T; and (iii) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can be any of A, C, G, or T, to produce polymerase chain reaction amplified fragments; and

20 (7) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands representing the 3'-ends of mRNAs present in the sample.

25 A depiction of this scheme is shown in Figure 1.

A. Isolation of mRNA

30 The first step in the method is isolation or provision of a mRNA population. Methods of extraction of RNA are well-known in the art and are described, for example, in J. Sambrook et al., "Molecular Cloning: A Laboratory Manual" (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989), vol. 1, ch. 7.

reference. Other isolation and extraction methods are

-20-

also well-known. Typically, isolation is performed in the presence of chaotropic agents such as guanidinium chloride or guanidinium thiocyanate, although other detergents and extraction agents can alternatively be
5 used.

Typically, the mRNA is isolated from the total extracted RNA by chromatography over oligo(dT)-cellulose or other chromatographic media that have the capacity to
10 bind the polyadenylated 3'-portion of mRNA molecules. Alternatively, but less preferably, total RNA can be used. However, it is generally preferred to isolate poly(A)⁺ RNA.

15 B. Preparation of Double-Stranded cDNA

Double-stranded cDNAs are then prepared from the mRNA population using a mixture of twelve anchor primers to initiate reverse transcription. The anchor primers each include: (i) a tract of from 7 to 40 T residues; (ii) a site for cleavage by a restriction endonuclease that recognizes more than six bases, the site for cleavage being located to the 5'-side of the tract of T residues; (iii) a stuffer segment of from 4 to 40 nucleotides, the stuffer segment being located to the
20 5'-side of the site for cleavage by the restriction endonuclease; and (iv) phasing residues -V-N located at the 3' end of each of the anchor primers, wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from
25 the group consisting of A, C, G, and T. The mixture includes anchor primers containing all possibilities for V and N.

segment of the anchor primers is 14 residues in length..

-21-

A suitable sequence of the stuffer segment is A-A-C-T-G-G-A-A-G-A-A-T-T-C (SEQ ID NO: 1). Typically, the site for cleavage by a restriction endonuclease that recognizes more than six bases is the NotI cleavage site.

5 A preferred set of anchor primers has the sequence A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2).

One member of this mixture of twelve anchor
10 primers initiates synthesis at a fixed position at the 3'-end of all copies of each mRNA species in the sample, thereby defining a 3'-end point for each species.

This reaction is carried out under conditions
15 for the preparation of double-stranded cDNA from mRNA that are well-known in the art. Such techniques are described, for example, in Volume 2 of J. Sambrook et al., "Molecular Cloning: A Laboratory Manual", entitled "Construction and Analysis of cDNA Libraries."
20 Typically, reverse transcriptase from avian myeloblastosis virus is used.

C. Cleavage of the cDNA Sample With Restriction Endonucleases

25 The cDNA sample is cleaved with two restriction endonucleases. The first restriction endonuclease is an endonuclease that recognizes a 4-nucleotide sequence. This typically cleaves at multiple sites in most cDNAs. The second restriction endonuclease cleaves at a single
30 site within each member of the mixture of anchor primers. Typically, the first restriction endonuclease is MspI and the second restriction endonuclease is NotI. The enzyme Not does not cleave within most cDNAs. This is desirable

anchor site.

-22-

Alternatively, the first restriction endonuclease can be TaqI or HinP1I. The use of the latter two restriction endonucleases can detect rare mRNAs that are not cleaved by MspI. The first 5 restriction endonuclease generates a 5'-overhang compatible for cloning into the desired vector, as discussed below. This cloning, for the pBC SK⁺ vector, is into the ClaI site, as discussed below.

10 Conditions for digestion of the cDNA are well-known in the art and are described, for example, in J. Sambrook et al., "Molecular Cloning: A Laboratory Manual," Vol. 1, Ch. 5, "Enzymes Used in Molecular Cloning."

15

D. Insertion of Cleaved cDNA into a Vector

The cDNA sample cleaved with the first and second restriction endonucleases is then inserted into a vector. A suitable vector is the plasmid pBC SK⁺ that has 20 been cleaved with the restriction endonucleases ClaI and NotI. The vector contains a bacteriophage-specific promoter. Typically, the promoter is a T3 promoter or a T7 promoter. A preferred promoter is bacteriophage T3 promoter. The cleaved cDNA is inserted into the promoter 25 in an orientation that is antisense with respect to the bacteriophage-specific promoter.

E. Transformation of a Suitable Host Cell

The vector into which the cleaved DNA has been 30 inserted is then used to transform a suitable host cell that can be efficiently transformed or transfected by the vector containing the insert. Suitable host cells for cloning are described, for example, in Sambrook et al.,

suitable host: Escherichia coli or Escherichia coli K12. A suitable host

5 ~~coli~~ strain is MC1061. Preferably, a small aliquot is also used to transform ~~E. coli~~ strain XL1-Blue so that the percentage of clones with inserts is determined from the relative percentages of blue and white colonies on X-gal plates. Only libraries with in excess of 5×10^5 recombinants are typically acceptable.

F. Generation of Linearized Fragments

10 Plasmid preparations, typically as minipreps, are then made from each of the cDNA libraries. Linearized fragments are then generated by digestion with at least one restriction endonuclease that is different from the first and second restriction endonucleases discussed above. Preferably, an aliquot of each of the 15 cloned inserts is divided into two pools, one of which is cleaved with XhoI and the second with SalI. The pools of linearized plasmids are combined, mixed, then divided into thirds. The thirds are digested with HindIII, BamHI, and EcoRI. This procedure is followed because, in 20 order to generate antisense transcripts of the inserts with T3 RNA polymerase, the template must first be cleaved with a restriction endonuclease that cuts within flanking sequences but not within the inserts themselves. Given that the average length of the 3'-terminal MspI 25 fragments is 256 base pairs, approximately 6% of the inserts contain sites for any enzyme with a hexamer recognition sequence. Those inserts would be lost to further analysis were only a single enzyme utilized. Hence, it is preferable to divide the reaction so that 30 only one of either of two enzymes is used for linearization of each half reaction. Only inserts containing sites for both enzymes (approximately 0.4%) are lost from both halves of the samples. Similarly,

multiple digestions can lead to incomplete cleavage and lead to variability in the efficiency of the after

-24-

polymerase chain reactions for different samples because
of differential competition for primers. Cleavage of
thirds of the samples with one of three enzymes that have
single targets in pBC SK⁺ between its ClaI and NotI sites
5 eliminates the production of transcripts containing
binding sites for the eventual 5' primers in the PCR
process from insertless plasmids. The use of three
enzymes on thirds of the reaction reduces the use of
insert-containing sequences that also contain sites for
10 the enzyme while solving the problem of possible
contamination of insertless sequences. If only one
enzyme were used, about 10% of the insert-containing
sequences would be lost, but this is reduced to about
0.1%, because only those sequences that fail to be
15 cleaved by all three enzymes are lost.

G. Generation of cRNA

The next step is a generation of a cRNA
preparation of antisense cRNA transcripts. This is
20 performed by incubation of the linearized fragments with
an RNA polymerase capable of initiating transcription
from the bacteriophage-specific promoter. Typically, as
discussed above, the promoter is a T3 promoter, and the
polymerase is therefore T3 RNA polymerase. The
25 polymerase is incubated with the linearized fragments and
the four ribonucleoside triphosphates under conditions
suitable for synthesis.

H. Transcription of First-Strand cDNA

30 The cRNA preparation is then divided into
sixteen subpools. First-strand cDNA is then transcribed
from each subpool, using a thermostable reverse
transcriptase and a primer as described below. A

available from Perkin-Elmer (Norwalk, CT). This enzyme

-25-

is also known as an RNA-dependent DNA polymerase. With
this reverse transcriptase, annealing is performed at
60°C, and the transcription reaction at 70°C. This
promotes high fidelity complementarity between the primer
5 and the cRNA. The primer used is one of the sixteen
primers whose 3'-terminus is -N-N, wherein N is one of
the four deoxyribonucleotides A, C, G, or T, the primer
being at least 15 nucleotides in length, corresponding in
sequence to the 3'-end of the bacteriophage-specific
10 promoter, and extending across into at least the first
two nucleotides of the cRNA.

Where the bacteriophage-specific promoter is
the T3 promoter, the primers typically have the sequence
15 A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3).

I. PCR Reaction

The next step is the use of the product of
transcription in each of the sixteen subpools as a
20 template for a polymerase chain reaction with primers as
described below to produce polymerase chain reaction
amplified fragments.

The primers used are: (a) a 3'-primer that
25 corresponds in sequence to a sequence in the vector
adjoining the site of insertion of the cDNA sample in the
vector; and (b) a 5'-primer selected from the group
consisting of: (i) the primer from which first-strand
cDNA was made for that subpool; (ii) the primer from
30 which the first-strand cDNA was made for that subpool
extended at its 3'-terminus by an additional residue -N,
where N can be any of A, C, G, or T; and (iii) the primer
used for the synthesis of first-strand cDNA for that

When the vector is the plasmid pBC SK⁺ cleaved with ClaI and NotI, a suitable 3'-primer is G-A-A-C-A-A-A-A-A-G-C-T-G-G-A-G-C-T-C-C-A-C-C-G-C (SEQ ID NO: 4). Where the bacteriophage-specific promoter is the T3 5 promoter, suitable 5'-primers have the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3), A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO: 5), or A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N-N (SEQ ID NO: 6).

10 Typically, PCR is performed in the presence of ³⁵S-dATP using a PCR program of 15 seconds at 94°C for denaturation, 15 seconds at 60°C for annealing, and 30 seconds at 72°C for synthesis on a Perkin-Elmer 9600 apparatus (Perkin-Elmer Cetus, Norwalk, CT). The high 15 temperature annealing step minimizes artifactual mispriming by the 5'-primer at its 3'-end and promotes high fidelity copying.

20 Alternatively, the PCR amplification can be carried out in the presence of a ³²P-labeled deoxyribonucleoside triphosphate, such as [³²P]dCTP. However, it is generally preferred to use a ³⁵S-labeled deoxyribonucleoside triphosphate for maximum resolution. Other detection methods, including nonradioactive labels, 25 can also be used.

These series of reactions produces 16, 64, and 256 product pools for the three sets of 5'-primers. It produces 16 product pools for the primer that is the same 30 as the primer from which first-strand cDNA was made. It produces 64 product pools for the primer extended at its 3'-terminus by an additional residue N, where N can be any of the four nucleotides. It produces 256 products

four nucleotides.

The process of the present invention can be extended by using longer sets of 5'-primers extended at their 3'-end by additional nucleotides. For example, a primer with the 3'-terminus -N-N-N-N-N would give 1024
5 products.

J. Electrophoresis

The polymerase chain reaction amplified fragments are then resolved by electrophoresis to display
10 bands representing the 3'-ends of mRNAs present in the sample.

Electrophoretic techniques for resolving PCR amplified fragments are well-understood in the art and
15 need not be further recited here. The corresponding products are resolved in denaturing DNA sequencing gels and visualized by autoradiography. For the particular vector system described herein, the gels are run so that the first 140 base pairs run off their bottom, since
20 vector-related sequences increase the length of the cDNAs by 140 base pairs. This number can vary if other vector systems are employed, and the appropriate electrophoresis conditions so that vector-related sequences run off the bottom of the gels can be determined from a consideration
25 of the sequences of the vector involved. Typically, each reaction is run on a separate denaturing gel, so that at least two gels are used. It is preferred to perform a series of reactions in parallel, such as from different tissues, and resolve all of the reactions using the same
30 primer on the same gel. A substantial number of reactions can be resolved on the same gel. Typically, as many as thirty reactions can be resolved on the same gel and compared. As discussed below, this provides a way of

Typically, autoradiography is used to detect the resolved cDNA species. However, other detection methods, such as phosphorimaging or fluorescence, can also be used, and may provide higher sensitivity in
5 certain applications.

According to the scheme, the cDNA libraries produced from each of the mRNA samples contain copies of the extreme 3'-ends from the most distal site for MspI to
10 the beginning of the poly(A) tail of all poly(A)⁺ mRNAs in the starting RNA sample approximately according to the initial relative concentrations of the mRNAs. Because both ends of the inserts for each species are exactly defined by sequence, their lengths are uniform for each
15 species allowing their later visualization as discrete bands on a gel, regardless of the tissue source of the mRNA.

The use of successive steps with lengthening
20 primers to survey the cDNAs essentially act like a nested PCR. These steps enhance quality control and diminish the background that potentially could result from amplification of untargeted cDNAs. In a preferred embodiment, the second reverse transcription step
25 subdivides each cRNA sample into sixteen subpools, utilizing a primer that anneals to the sequences derived from pBC SK⁺ but extends across the CGG of the non-regenerated MspI site and including two nucleotides (-N-N) of the insert. This step segregates the starting
30 population of potentially 50,000 to 100,000 mRNAs into sixteen subpools of approximately 3,000 to 6,000 members each. In serial iterations of the subsequent PCR step, in which radioactive label is incorporated into the

subdivided subsubpools by using progressively longer 3'

-29-

primers containing three or four nucleotides of the insert.

By first demanding by high temperature
5 annealing a high fidelity 3'-end match at the reverse transcription step in the -N-N positions, and subsequently demanding again such high fidelity matching into -N-N-N or -N-N-N-N iterations, bleedthrough from mismatched priming at the -N-N positions is drastically
10 minimized.

The steps of the process beginning with dividing the cRNA preparation into sixteen subpools and transcribing first-strand cDNA from each subpool can be
15 performed separately as a method of simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs.

20 II. APPLICATIONS OF THE METHOD FOR DISPLAY OF mRNA PATTERNS

The method described above for the detection of patterns of mRNA expression in a tissue and the resolving of these patterns by gel electrophoresis has a number of
25 applications. One of these applications is its use for the detection of a change in the pattern of mRNA expression in a tissue associated with a physiological or pathological change. In general, this method comprises:

(1) obtaining a first sample of a tissue that
30 is not subject to the physiological or pathological change;

(2) determining the pattern of mRNA expression in the first sample of the tissue by performing the

pool representing the 3'-ends of a population of mRNAs as

-30-

described above to generate a first display of bands representing the 3'-ends of mRNAs present in the first sample;

5 (3) obtaining a second sample of the tissue that has been subject to the physiological or pathological change;

10 (4) determining the pattern of mRNA expression in the second sample of the tissue by performing the method of simultaneous sequence-specific identification 10 of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs as described above to generate a second display of bands representing the 3'-ends of mRNAs present in the second sample; and

15 (5) comparing the first and second displays to determine the effect of the physiological or pathological change on the pattern of mRNA expression in the tissue.

Typically, the comparison is made in adjacent 20 lanes of a single gel.

The tissue can be derived from the central nervous system. In particular, it can be derived from a structure within the central nervous system that is the 25 retina, cerebral cortex, olfactory bulb, thalamus, hypothalamus, anterior pituitary, posterior pituitary, hippocampus, nucleus accumbens, amygdala, striatum, cerebellum, brain stem, suprachiasmatic nucleus, or spinal cord. When the tissue is derived from the central 30 nervous system, the physiological or pathological change can be any of Alzheimer's disease, parkinsonism, ischemia, alcohol addiction, drug addiction, schizophrenia, amyotrophic lateral sclerosis, multiple

aging, or long-term potentiation, the latter affecting the hippocampus. Additionally, particularly with reference to mRNA species occurring in particular structures within the central nervous system, the method 5 can be used to study brain regions that are known to be involved in complex behaviors, such as learning and memory, emotion, drug addiction, glutamate neurotoxicity, feeding behavior, olfaction, viral infection, vision, and movement disorders.

10

This method can also be used to study the results of the administration of drugs and/or toxins to an individual by comparing the mRNA pattern of a tissue before and after the administration of the drug or toxin. 15 Results of electroshock therapy can also be studied.

Alternatively, the tissue can be from an organ or organ system that includes the cardiovascular system, the pulmonary system, the digestive system, the 20 peripheral nervous system, the liver, the kidney, skeletal muscle, and the reproductive system, or from any other organ or organ system of the body. For example, mRNA patterns can be studied from liver, heart, kidney, or skeletal muscle. Additionally, for any tissue, 25 samples can be taken at various times so as to discover a circadian effect of mRNA expression. Thus, this method can ascribe particular mRNA species to involvement in particular patterns of function or malfunction.

30

The antisense cRNA pool representing the 3'-ends of mRNAs can be generated by steps (1)-(4) of the method as described above in Section I.

screening for a side effect of a drug. In general, such a method comprises:

- (1) obtaining a first sample of tissue from an organism treated with a compound of known physiological function;
- 5 (2) determining the pattern of mRNA expression in the first sample of the tissue by performing the method of simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs, as described above, to generate a first display of bands representing the 3'-ends of mRNAs present in the first sample;
- 10 (3) obtaining a second sample of tissue from an organism treated with a drug to be screened for a side effect;
- 15 (4) determining the pattern of mRNA expression in the second sample of the tissue by performing the method of simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs, as described above, to generate a second display of bands representing the 3'-ends of mRNAs present in the second sample; and
- 20 (5) comparing the first and second displays in order to detect the presence of mRNA species whose expression is not affected by the known compound but is affected by the drug to be screened, thereby indicating a difference in action of the drug to be screened and the known compound and thus a side effect.
- 25
- 30

In particular, this method can be used for drugs affecting the central nervous system, such as

stimulants. However, this method can in fact be used for

-33-

any drug that may affect mRNA expression in a particular tissue. For example, the effect on mRNA expression of anti-parkinsonism agents, skeletal muscle relaxants, analgesics, local anesthetics, cholinergics,
5 antispasmodics, steroids, non-steroidal anti-inflammatory drugs, antiviral agents, or any other drug capable of affecting mRNA expression can be studied, and the effect determined in a particular tissue or structure.

10 A further application of the method of the present invention is in obtaining the sequence of the 3'-ends of mRNA species that are displayed. In general, a method of obtaining the sequence comprises:

- 15 (1) eluting at least one cDNA corresponding to a mRNA from an electropherogram in which bands representing the 3'-ends of mRNAs present in the sample are displayed;
- (2) amplifying the eluted cDNA in a polymerase chain reaction;
- 20 (3) cloning the amplified cDNA into a plasmid;
- (4) producing DNA corresponding to the cloned DNA from the plasmid; and
- (5) sequencing the cloned cDNA.

25 The cDNA that has been excised can be amplified with the primers previously used in the PCR step. The cDNA can then be cloned into pCR II (Invitrogen, San Diego, CA) by TA cloning and ligation into the vector. Minipreps of the DNA can then be produced by standard
30 techniques from subclones and a portion denatured and split into two aliquots for automated sequencing by the dideoxy chain termination method of Sanger. A commercially available sequencer can be used, such as a

cDNAs studied, in the length range of 50-500 bp, across the entire length of the fragment.

These partial sequences can then be used to
5 scan genomic data bases such as GenBank to recognize sequence identities and similarities using programs such as BLASTN and BLASTX. Because this method generates sequences from only the 3'-ends of mRNAs it is expected that open reading frames (ORFs) would be encountered only
10 occasionally, as the 3'-untranslated regions of brain mRNAs are on average longer than 1300 nucleotides (J.G. Sutcliffe, supra). Potential ORFs can be examined for signature protein motifs.

15 The cDNA sequences obtained can then be used to design primer pairs for semiquantitative PCR to confirm tissue expression patterns. Selected products can also be used to isolate full-length cDNA clones for further analysis. Primer pairs can be used for SSCP-PCR (single
20 strand conformation polymorphism-PCR) amplification of genomic DNA. For example, such amplification can be carried out from a panel of interspecific backcross mice to determine linkage of each PCR product to markers already linked. This can result in the mapping of new
25 genes and can serve as a resource for identifying candidates for mapped mouse mutant loci and homologous human disease genes. SSCP-PCR uses synthetic oligonucleotide primers that amplify, via PCR, a small (100-200 bp) segment. (M. Orita et al., "Detection of
30 Polymorphisms of Human DNA by Gel Electrophoresis as Single-Strand Conformation Polymorphisms," Proc. Natl. Acad. Sci. USA 86: 2766-2770 (1989); M. Orita et al., "Rapid and Sensitive Detection of Point Mutations in DNA

The excised fragments of cDNA can be radiolabeled by techniques well-known in the art for use in probing a northern blot or for in situ hybridization to verify mRNA distribution and to learn the size and 5 prevalence of the corresponding full-length mRNA. The probe can also be used to screen a cDNA library to isolate clones for more reliable and complete sequence determination. The labeled probes can also be used for any other purpose, such as studying in vitro expression.

10

III. PANELS AND DEGENERATE MIXTURES OF PRIMERS

Another aspect of the present invention is panels and degenerate mixtures of primers suitable for 15 the practice of the present invention. These include:

- (1) a panel of primers comprising 16 primers of the sequence A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3), wherein N is one of the four deoxyribonucleotides A, C, G, or T;
- 20 (2) a panel of primers comprising 64 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO: 5), wherein N is one of the four deoxyribonucleotides A, C, G, or T;
- (3) a panel of primers comprising 256 primers 25 of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N-N (SEQ ID NO: 6), wherein N is one of the four deoxyribonucleotides A, C, G, or T; and
- (4) a panel of primers comprising 12 primers 30 of the sequences A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group

is a degenerate mixture of primers comprising a mixture of 12 primers of the sequences A-A-C-T-G-G-A-A-

G-A-A-T-T-C-G-C-G-C-C-G-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from 5 the group consisting of A, C, G, and T, each of the 12 primers being present in about an equimolar quantity.

The invention is illustrated by the following Example. The Example is for illustrative purposes only 10 and is not intended to limit the invention.

EXAMPLE

15 Resolution of Brain mRNAs Using Primers Corresponding to
Sequences of Known Brain mRNAs of Different
Concentrations

To demonstrate the effectiveness of the method of the present invention, it was applied using 5'-primers 20 extended at their 3'-ends by two nucleotides and corresponding to the sequence of known brain mRNAs of different concentrations, such as neuron-specific enolase (NSE) at roughly 0.5% concentration (S. Forss-Petter et al., "Neuron-Specific Enolase: Complete Structure of Rat 25 mRNA, Multiple Transcriptional Start Sites and Evidence for Translational Control," J. Neurosci. Res. 16: 141-156 (1986)), RC3 at about 0.01%, and somatostatin at 0.001% (G.H. Travis & J.G. Sutcliffe, "Phenol Emulsion-Enhanced DNA-Driven Subtractive cDNA Cloning: Isolation of Low- 30 Abundance Monkey Cortex-Specific mRNAs," Proc. Natl. Acad. Sci. USA 85: 1696-1700 (1988)) to compare cDNAs made from libraries constructed from cerebral cortex, striatum, cerebellum and liver RNAs made as described

were absolutely reproducible in duplicate samples.

-37-

Approximately two-thirds of the bands differed between brain and liver samples, including the bands of the correct lengths corresponding to the known brain-specific mRNAs. This was confirmed by excision of the bands from 5 the gels, amplification and sequencing. Only a few bands differed among samples for various brain regions for any particular primer, although some band intensities differed.

10 The band corresponding to NSE, a relatively prevalent mRNA species, appeared in all of the brain samples but not in the liver samples, but was not observed when any of the last three single nucleotides within the four-base 3'-terminal sequence -N-N-N-N was 15 changed in the synthetic 5'-primer. When the first N was changed, a small amount of bleedthrough is detected. For the known species, the intensity of the autoradiographic signal was roughly proportional to mRNA prevalence, and mRNAs with concentrations of one part in 10^5 or greater of 20 the poly(A)⁺ RNA were routinely visible, with the occasional problem that cDNAs that migrated close to more intense bands were obscured.

A sample of the data is shown in Figure 2. In 25 the 5 gel lanes on the left, cortex cRNA was substrate for reverse transcription with the primer A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3) where -N-N is -C-T (primer 118), -G-T (primer 116) or -C-G (primer 106). The PCR amplification used primers A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N-N (SEQ ID NO: 6) where -N-N-N-N is -C-T-A-C (primer 128), -C-T-G-A (primer 127), -C-T-G-C (primer 111), -G-T-G-C (primer 134), and -C-G-G-C (primer 130), 30 as indicated in Figure 2. Primers 118 and 111 match the

downstream from the NspI site located the nearest to the end of the NSE mRNA sequence. Primer 127 is mismatched

-38-

with the NSE sequence in the last (-1) position, primer 128 in the next-to-last (-2) position, primers 106 and 130 in the -3 position, and primers 116 and 134 in the -4 position. Primer 134 extended two nucleotides further upstream than the others shown here, hence its PCR products are two nucleotides longer relative to the products in other lanes.

In each lane, 50-100 bands were visible in 15-minute exposures using ^{32}P -dCTP to radiolabel the products. These bands were apparently distinct for each primer pair, with the exception that a subset of the 118-111 bands appeared more faintly in the 116-134 lane, trailing by two nucleotides, indicating bleedthrough in the four position.

The 118-111 primer set was used again on separate cortex (CX) and liver (LV) cRNAs. The cortex pattern was identical to that in lane 118-111, demonstrating reproducibility. The liver pattern differed from CX in the majority of species. The asterisk indicates the position of the NSE product. Analogous primer sets detected RC3 and somatostatin (somat) products (asterisks) in CX but not LV lanes. The relative band intensities of a given PCR product can be compared within lanes using the same primer set, but not different sets.

This example demonstrates the feasibility and reproducibility of the method of the present invention and its ability to resolve different mRNAs. It further demonstrates that prevalence of particular mRNA species can be estimated from the intensity of the

... from right and left panels
simultaneously.

ADVANTAGES OF THE PRESENT INVENTION

The present method can be used to identify genes whose expression is altered during neuronal development, in models of plasticity and regeneration, in 5 response to chemical or electrophysiological challenges such as neurotoxicity and long-term potentiation, and in response to behavioral, viral, drug/alcohol paradigms, the occurrence of cell death or apoptosis, aging, pathological conditions, and other conditions affecting 10 mRNA expression. Although the method is particularly useful for studying gene expression in the nervous system, it is not limited to the nervous system and can be used to study mRNA expression in any tissue. The method allows the visualization of nearly every mRNA expressed 15 by a tissue as a distinct band on a gel whose intensity corresponds roughly to the concentration of the mRNA.

The method has the advantage that it does not depend on potentially irreproducible mismatched random 20 priming, so that it provides a high degree of accuracy and reproducibility. Moreover, it reduces the complications and imprecision generated by the presence of concurrent bands of different length resulting from the same mRNA species as the result of different priming 25 events. In methods using random priming, such concurrent bands can occur and are more likely to occur for mRNA species of high prevalence. In the present method, such concurrent bands are avoided.

30 The method provides sequence-specific information about the mRNA species and can be used to generate primers, probes, and other specific sequences.

DETAILED DESCRIPTION OF THE INVENTION

As described in considerable detail, with reference to certain preferred versions thereof, other versions are

- 4 0 -

possible. Therefore, the spirit and scope of the appended claims should not be limited to the description of the preferred versions contained herein.

-41-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: The Scripps Research Institute
 - (B) STREET: 10666 North Torrey Pines Road
 - (C) CITY: La Jolla
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) POSTAL CODE (ZIP): 92037
 - (G) TELEPHONE: (619) 455-9100
 - (H) TELEFAX: (619) 554-6612
-
- (A) NAME: ERLANDER, Mark G.
 - (B) STREET: 1352 Via Terrassa
 - (C) CITY: Encinitas
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) POSTAL CODE (ZIP): 92024
-
- (A) NAME: SUTCLIFFE, Gregor J.
 - (B) STREET: 2253 Via Tiempo
 - (C) CITY: Cardiff
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) POSTAL CODE (ZIP): 92007

(ii) TITLE OF INVENTION: Method for Simultaneous Identification of Differentially Expressed mRNAs and Measurement of Relative Concentrations

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(v) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: PCT US94/
- (B) FILING DATE: 14-NOV-1994
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/152,482
- (B) FILING DATE: 12-NOV-1993
- (C) CLASSIFICATION: 435

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs

-42-

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AACTGGAAGA ATTC

14

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AACTGGAAGA ATT CGCGGCC GCAGGAATTT TTTTTTTTTT TTTTTVN

47

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- 43 -

(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGGTCGACGG TATCGGNN

18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAACAAAAGC TGGAGCTCCA CCGC

24

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGGTCGACGG TATCGNNNN

19

A LENGTH: 19 base pairs
(B) TYPE: nucleic acid

- 44 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Synthetic primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGGTCGACGG TATCGGNNNN

20

We claim:

1. A method for simultaneous sequence-specific identification of mRNAs in a mRNA population comprising the steps of:
 - (a) preparing double-stranded cDNAs from a mRNA population using a mixture of 12 anchor primers, the anchor primers each including: (i) a tract of from 7 to 40 T residues; (ii) a site for cleavage by a restriction endonuclease that recognizes more than six bases, the site for cleavage being located to the 5'-side of the tract of T residues; (iii) a stuffer segment of from 4 to 40 nucleotides, the stuffer segment being located to the 5'-side of the site for cleavage by the restriction endonuclease; and (iv) phasing residues -V-N located at the 3' end of each of the anchor primers, wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T, the mixture including anchor primers containing all possibilities for V and N;
 - (b) producing cloned inserts from a suitable host cell that has been transformed by a vector, the vector having the cDNA sample that has been cleaved with a first restriction endonuclease and a second restriction endonuclease inserted therein, the cleaved cDNA sample being inserted in the vector in an orientation that is antisense with respect to a bacteriophage-specific promoter within the vector, the first restriction endonuclease recognizing a four-nucleotide sequence and the second restriction endonuclease cleaving at a single site within each member of the mixture of anchor primers;
 - (c) generating linearized fragments of the

restriction endonucleases;

- (d) generating a cRNA preparation of antisense cRNA transcripts by incubation of the linearized fragments with a bacteriophage-specific RNA polymerase capable of initiating transcription from the 5 bacteriophage-specific promoter;
- (e) dividing the cRNA preparation into sixteen subpools and transcribing first-strand cDNA from each subpool, using a thermostable reverse transcriptase and one of sixteen primers whose 3'-terminus is -N-N, wherein 10 N is one of the four deoxyribonucleotides A, C, G, or T, the primer being at least 15 nucleotides in length, corresponding in sequence to the 3'-end of the bacteriophage-specific promoter, and extending across 15 into at least the first two nucleotides of the cRNA, the mixture including all possibilities for the 3'-terminal two nucleotides;
- (f) using the product of transcription in each of the sixteen subpools as a template for a polymerase chain reaction with a 3'-primer that corresponds in 20 sequence to a sequence in the vector adjoining the site of insertion of the cDNA sample in the vector and a 5'-primer selected from the group consisting of: (i) the primer from which first-strand cDNA was made for that subpool; (ii) the primer from which the first-strand cDNA 25 was made for that subpool extended at its 3'-terminus by an additional residue -N, where N can be any of A, C, G, or T; and (iii) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can 30 be any of A, C, G, or T, to produce polymerase chain reaction amplified fragments; and
- (g) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands

- 47 -

2. The method of claim 1 wherein the anchor primers each have 18 T residues in the tract of T residues.

5 3. The method of claim 1 wherein the stuffer segment of the anchor primers is 14 residues in length.

4. The method of claim 3 wherein the sequence of the stuffer segment is A-A-C-T-G-G-A-A-G-A-A-T-T-C
10 (SEQ ID NO: 1).

5. The method of claim 1 wherein the site for cleavage by a restriction endonuclease that recognizes more than six bases is the NotI cleavage site.

15 6. The method of claim 4 wherein the anchor primers have the sequence A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2).

20 7. The method of claim 1 wherein the bacteriophage-specific promoter is selected from the group consisting of T3 promoter and T7 promoter.

25 8. The method of claim 7 wherein the bacteriophage-specific promoter is T3 promoter.

9. The method of claim 8 wherein the sixteen primers for priming of transcription of cDNA from cRNA
30 have the sequence A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3).

10. The method of claim 1 wherein the vector
T-C-G-A-U-C-G-C (SEQ ID NO: 4).

- 48 -

11. The method of claim 1 wherein the first restriction endonuclease recognizing a four-nucleotide sequence is MspI.

5 12. The method of claim 1 wherein the first restriction endonuclease recognizing a four-nucleotide sequence is selected from the group consisting of TagI and HinP1I.

10 13. The method of claim 1 wherein the restriction endonuclease cleaving at a single site in each of the mixture of anchor primers is NotI.

15 14. The method of claim 1 wherein the step of generating linearized fragments of the cloned inserts comprises:

20 (i) dividing the plasmid containing the insert into two fractions, a first fraction cleaved with the restriction endonuclease XhoI and a second fraction cleaved with the restriction endonuclease SalI;

(ii) recombining the first and second fractions after cleavage;

25 (iii) dividing the recombined fractions into thirds and cleaving the first third with the restriction endonuclease HindIII, the second third with the restriction endonuclease BamHI, and the third third with the restriction endonuclease EcoRI; and

30 (iv) recombining the thirds after digestion in order to produce a population of linearized fragments of which about one-sixth of the population corresponds to the product of cleavage by each of the possible combinations of enzymes.

species.

-49-

16. The method of claim 1 wherein the intensity of each band displayed after electrophoresis is about proportional to the abundance of the mRNA corresponding to the band in the original mixture.

5

17. The method of claim 16 further comprising a step of determining the relative abundance of each mRNA in the original mixture from the intensity of the band corresponding to that mRNA after electrophoresis.

10

18. The method of claim 1 wherein the step of resolving the polymerase chain reaction amplified fragments by electrophoresis comprises electrophoresis of the fragments on at least two gels.

15

19. The method of claim 1 wherein the suitable host cell is Escherichia coli.

20. The method of claim 1 further comprising the steps of:

(h) eluting at least one cDNA corresponding to a mRNA from an electropherogram in which bands representing the 3'-ends of mRNAs present in the sample are displayed;

25 (i) amplifying the eluted cDNA in a polymerase chain reaction;

(j) cloning the amplified cDNA into a plasmid;

(k) producing DNA corresponding to the cloned DNA from the plasmid; and

30 (l) sequencing the cloned cDNA.

21. A method for simultaneous sequence-specific identification of mRNAs in a mRNA population

5 (b) preparing double-stranded cDNAs from the mRNA population using a mixture of 12 anchor primers with the sequence A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T, the mixture including anchor primers containing all possibilities for V and N, to produce a

10 cDNA sample;

15 (c) cleaving the cDNA sample with two restriction endonucleases, a first restriction endonuclease MspI and a second restriction endonuclease NotI;

20 (d) inserting the cDNA sample cleaved with the first and second restriction endonucleases into a vector, the cleaved cDNA being inserted in an orientation that is antisense with respect to a T3 promoter within the vector, the vector being the plasmid pBC SK^{*} cleaved with ClaI and NotI;

25 (e) transforming Escherichia coli with the vector into which the cleaved cDNA has been inserted to produce cloned inserts;

30 (f) generating linearized fragments of the cloned inserts by digestion with at least one restriction endonuclease that is different from the first and second restriction endonucleases;

(g) generating a cRNA preparation of antisense cRNA transcripts by incubation of the linearized fragments with a T3 RNA polymerase capable of initiating transcription from the T3-specific promoter;

(h) dividing the cRNA preparation into sixteen subpools and transcribing first-strand cDNA from each

deoxyribonucleotides A, C, G, or T, the mixture including all possibilities for the 3'-terminal two nucleotides;

(i) using the product of transcription in each of the sixteen subpools as a template for a polymerase chain reaction with the 3'-primer G-A-A-C-A-A-A-G-C-T-G-G-A-G-C-T-C-C-A-C-C-G-C (SEQ ID NO: 4), and a 5'-primer selected from the group consisting of: (1) the primer from which first-strand cDNA was made for that subpool; (2) the primer from which the first-strand cDNA was made for that subpool extended at its 3'-terminus by an additional residue -N, where N can be any of A, C, G, or T; and (3) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can be any of A, C, G, or T, to produce polymerase chain reaction amplified fragments; and

(j) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands representing the 3'-ends of mRNAs present in the sample.

20

22. A method of simultaneous sequence-specific identification of mRNAs corresponding to members of an antisense cRNA pool representing the 3'-ends of a population of mRNAs, the antisense cRNAs that are members of the antisense cRNA pool being terminated at their 5'-end with a primer sequence corresponding to a bacteriophage-specific vector and at their 3'-end with a sequence corresponding in sequence to a sequence of the vector, the method comprising;

30 (a) dividing the members of the antisense cRNA pool into sixteen subpools and transcribing first-strand cDNA from each subpool, using a thermostable reverse transcriptase and one of sixteen primers, where the

10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400
4410
4420
4430
4440
4450
4460
4470
4480
4490
4500
4510
4520
4530
4540
4550
4560
4570
4580
4590
4600
4610
4620
4630
4640
4650
4660
4670
4680
4690
4700
4710
4720
4730
4740
4750
4760
4770
4780
4790
4800
4810
4820
4830
4840
4850
4860
4870
4880
4890
4900
4910
4920
4930
4940
4950
4960
4970
4980
4990
5000
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
5710
5720
5730
5740
5750
5760
5770
5780
5790
5800
5810
5820
5830
5840
5850
5860
5870
5880
5890
5900
5910
5920
5930
5940
5950
5960
5970
5980
5990
6000
6010
6020
6030
6040
6050
6060
6070
6080
6090
6100
6110
6120
6130
6140
6150
6160
6170
6180
6190
6200
6210
6220
6230
6240
6250
6260
6270
6280
6290
6300
6310
6320
6330
6340
6350
6360
6370
6380
6390
6400
6410
6420
6430
6440
6450
6460
6470
6480
6490
6500
6510
6520
6530
6540
6550
6560
6570
6580
6590
6600
6610
6620
6630
6640
6650
6660
6670
6680
6690
6700
6710
6720
6730
6740
6750
6760
6770
6780
6790
6800
6810
6820
6830
6840
6850
6860
6870
6880
6890
6900
6910
6920
6930
6940
6950
6960
6970
6980
6990
7000
7010
7020
7030
7040
7050
7060
7070
7080
7090
7100
7110
7120
7130
7140
7150
7160
7170
7180
7190
7200
7210
7220
7230
7240
7250
7260
7270
7280
7290
7300
7310
7320
7330
7340
7350
7360
7370
7380
7390
7400
7410
7420
7430
7440
7450
7460
7470
7480
7490
7500
7510
7520
7530
7540
7550
7560
7570
7580
7590
7600
7610
7620
7630
7640
7650
7660
7670
7680
7690
7700
7710
7720
7730
7740
7750
7760
7770
7780
7790
7800
7810
7820
7830
7840
7850
7860
7870
7880
7890
7900
7910
7920
7930
7940
7950
7960
7970
7980
7990
8000
8010
8020
8030
8040
8050
8060
8070
8080
8090
8100
8110
8120
8130
8140
8150
8160
8170
8180
8190
8200
8210
8220
8230
8240
8250
8260
8270
8280
8290
8300
8310
8320
8330
8340
8350
8360
8370
8380
8390
8400
8410
8420
8430
8440
8450
8460
8470
8480
8490
8500
8510
8520
8530
8540
8550
8560
8570
8580
8590
8600
8610
8620
8630
8640
8650
8660
8670
8680
8690
8700
8710
8720
8730
8740
8750
8760
8770
8780
8790
8800
8810
8820
8830
8840
8850
8860
8870
8880
8890
8900
8910
8920
8930
8940
8950
8960
8970
8980
8990
9000
9010
9020
9030
9040
9050
9060
9070
9080
9090
9100
9110
9120
9130
9140
9150
9160
9170
9180
9190
9200
9210
9220
9230
9240
9250
9260
9270
9280
9290
9300
9310
9320
9330
9340
9350
9360
9370
9380
9390
9400
9410
9420
9430
9440
9450
9460
9470
9480
9490
9500
9510
9520
9530
9540
9550
9560
9570
9580
9590
9600
9610
9620
9630
9640
9650
9660
9670
9680
9690
9700
9710
9720
9730
9740
9750
9760
9770
9780
9790
9800
9810
9820
9830
9840
9850
9860
9870
9880
9890
9900
9910
9920
9930
9940
9950
9960
9970
9980
9990
10000
10010
10020
10030
10040
10050
10060
10070
10080
10090
100100
100200
100300
100400
100500
100600
100700
100800
100900
1001000
1002000
1003000
1004000
1005000
1006000
1007000
1008000
1009000
10010000
10020000
10030000
10040000
10050000
10060000
10070000
10080000
10090000
100100000
100200000
100300000
100400000
100500000
100600000
100700000
100800000
100900000
1001000000
1002000000
1003000000
1004000000
1005000000
1006000000
1007000000
1008000000
1009000000
10010000000
10020000000
10030000000
10040000000
10050000000
10060000000
10070000000
10080000000
10090000000
100100000000
100200000000
100300000000
100400000000
100500000000
100600000000
100700000000
100800000000
100900000000
1001000000000
1002000000000
1003000000000
1004000000000
1005000000000
1006000000000
1007000000000
1008000000000
1009000000000
10010000000000
10020000000000
10030000000000
10040000000000
10050000000000
10060000000000
10070000000000
10080000000000
10090000000000
100100000000000
100200000000000
100300000000000
100400000000000
100500000000000
100600000000000
100700000000000
100800000000000
100900000000000
1001000000000000
1002000000000000
1003000000000000
1004000000000000
1005000000000000
1006000000000000
1007000000000000
1008000000000000
1009000000000000
10010000000000000
10020000000000000
10030000000000000
10040000000000000
10050000000000000
10060000000000000
10070000000000000
10080000000000000
10090000000000000
100100000000000000
100200000000000000
100300000000000000
100400000000000000
100500000000000000
100600000000000000
100700000000000000
100800000000000000
100900000000000000
1001000000000000000
1002000000000000000
1003000000000000000
1004000000000000000
1005000000000000000
1006000000000000000
1007000000000000000
1008000000000000000
1009000000000000000
10010000000000000000
10020000000000000000
10030000000000000000
10040000000000000000
10050000000000000000
10060000000000000000
10070000000000000000
10080000000000000000
10090000000000000000
100100000000000000000
100200000000000000000
100300000000000000000
100400000000000000000
100500000000000000000
100600000000000000000
100700000000000000000
100800000000000000000
100900000000000000000
1001000000000000000000
1002000000000000000000
1003000000000000000000
1004000000000000000000
1005000000000000000000
1006000000000000000000
1007000000000000000000
1008000000000000000000
1009000000000000000000
10010000000000000000000
10020000000000000000000
10030000000000000000000
10040000000000000000000
10050000000000000000000
10060000000000000000000
10070000000000000000000
10080000000000000000000
10090000000000000000000
100100000000000000000000
100200000000000000000000
100300000000000000000000
100400000000000000000000
100500000000000000000000
100600000000000000000000
100700000000000000000000
100800000000000000000000
100900000000000000000000
1001000000000000000000000
1002000000000000000000000
1003000000000000000000000
1004000000000000000000000
1005000000000000000000000
1006000000000000000000000
1007000000000000000000000
1008000000000000000000000
1009000000000000000000000
10010000000000000000000000
10020000000000000000000000
10030000000000000000000000
10040000000000000000000000
10050000000000000000000000
10060000000000000000000000
10070000000000000000000000
10080000000000000000000000
10090000000000000000000000
100100000000000000000000000
100200000000000000000000000
100300000000000000000000000
100400000000000000000000000
100500000000000000000000000
100600000000000000000000000
100700000000000000000000000
100800000000000000000000000
100900000000000000000000000
1001000000000000000000000000
1002000000000000000000000000
1003000000000000000000000000
1004000000000000000000000000
1005000000000000000000000000
1006000000000000000000000000
1007000000000000000000000000
1008000000000000000000000000
1009000000000000000000000000
10010000000000000000000000000
10020000000000000000000000000
10030000000000000000000000000
10040000000000000000000000000
10050000000000000000000000000
10060000000000000000000000000
10070000000000000000000000000
10080000000000000000000000000
10090000000000000000000000000
100100000000000000000000000000
100200000000000000000000000000
100300000000000000000000000000
100400000000000000000000000000
100500000000000000000000000000
100600000000000000000000000000
100700000000000000000000000000
100800000000000000000000000000
100900000000000000000000000000
1001000000000000000000000000000
1002000000000000000000000000000
1003000000000000000000000000000
1004000000000000000000000000000
1005000000000000000000000000000
10060000000000000000

to the 3'-end of the bacteriophage-specific promoter, and extending across into at least the first two nucleotides of the cRNA, the mixture including all possibilities for the 3'-terminal two nucleotides;

- 5 (b) using the product of transcription in each of the sixteen subpools as a template for a polymerase chain reaction with a 3'-primer that corresponds in sequence to a sequence vector adjoining the site of insertion of the cDNA sample in the vector and a 5'-primer selected from the group consisting of: (i) the primer from which first-strand cDNA was made for that subpool; (ii) the primer from which the first-strand cDNA was made for that subpool extended at its 3'-terminus by an additional residue -N, where N can be any of A, C, G, or T; and (iii) the primer used for the synthesis of first-strand cDNA for that subpool extended at its 3'-terminus by two additional residues -N-N, wherein N can be any of A, C, G, or T, to produce polymerase chain reaction amplified fragments; and
- 10 (c) resolving the polymerase chain reaction amplified fragments by electrophoresis to display bands representing the 3'-ends of mRNAs present in the sample.

- 25 23. A method for detecting a change in the pattern of mRNA expression in a tissue associated with a physiological or pathological change comprising the steps of:

- 30 (a) obtaining a first sample of a tissue that is not subject to the physiological or pathological change;
- (b) determining the pattern of mRNA expression in the first sample of the tissue by performing steps (a)-(c) of claim 22 to generate a first display of bands

-53-

(c) obtaining a second sample of the tissue that has been subject to the physiological or pathological change;

5 (d) determining the pattern of mRNA expression in the second sample of the tissue by performing steps (a)-(c) of claim 22 to generate a second display of bands representing the 3'-ends of mRNAs present in the second sample; and

10 (e) comparing the first and second displays to determine the effect of the physiological or pathological change on the pattern of mRNA expression in the tissue.

24. The method of claim 23 wherein the tissue is derived from the central nervous system.

15

25. The method of claim 24 wherein the physiological or pathological change is selected from the group consisting of Alzheimer's disease, parkinsonism, ischemia, alcohol addiction, drug addiction, schizophrenia, amyotrophic lateral sclerosis, multiple sclerosis, depression, and bipolar manic-depressive disorder.

26. The method of claim 24 wherein the physiological or pathological change is associated with learning or memory, emotion, glutamate neurotoxicity, feeding behavior, olfaction, vision, movement disorders, viral infection, electroshock therapy, or the administration of a drug or toxin.

30

27. The method of claim 24 wherein the physiological or pathological change is selected from the group consisting of circadian variation, aging, and long

28. The method of claim 24 wherein the tissue
is derived from a structure within the central nervous
system selected from the group consisting of retina,
cerebral cortex, olfactory bulb, thalamus, hypothalamus,
5 anterior pituitary, posterior pituitary, hippocampus,
nucleus accumbens, amygdala, striatum, cerebellum, brain
stem, suprachiasmatic nucleus, and spinal cord.

29. The method of claim 23 wherein the tissue
10 is from an organ or organ system selected from the group
consisting of the cardiovascular system, the pulmonary
system, the digestive system, the peripheral nervous
system, the liver, the kidney, skeletal muscle, and the
reproductive system.

15

30. A method of screening for a side effect of
a drug comprising the steps of:

(a) obtaining a first sample of tissue from an
organism treated with a compound of known physiological
20 function;

(b) determining the pattern of mRNA expression
in the first sample of the tissue by performing steps
(a)-(c) of claim 22 to generate a first display of bands
representing the 3'-ends of mRNAs present in the first
25 sample;

(c) obtaining a second sample of tissue from
an organism treated with a drug to be screened for a side
effect;

(d) determining the pattern of mRNA expression
30 in the second sample of the tissue by performing steps
(a)-(c) of claim 22 to generate a second display of bands
representing the 3'-ends of mRNAs present in the second
sample; and

expression is not affected by the known compound but is

-55-

affected by the drug to be screened, thereby indicating a difference in action of the drug to be screened and the known compound and thus a side effect.

5 31. The method of claim 30 wherein the drug to be tested is selected from the group consisting of antidepressants, neuroleptics, tranquilizers, anticonvulsants, monoamine oxidase inhibitors, and stimulants.

10 32. The method of claim 30 wherein the drug to be tested is selected from the group consisting of anti-parkinsonism agents, skeletal muscle relaxants, analgesics, local anesthetics, cholinergics, antiviral 15 agents, antispasmodics, steroids, and non-steroidal anti-inflammatory drugs.

20 33. A panel of primers comprising 16 primers of the sequence A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N (SEQ ID NO: 3), wherein N is one of the four deoxyribonucleotides A, C, G, or T.

25 34. A panel of primers comprising 64 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO: 5), wherein N is one of the four deoxyribonucleotides A, C, G, or T.

30 35. A panel of primers comprising 256 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-N-N-N (SEQ ID NO: 6), wherein N is one of the four deoxyribonucleotides A, C, G, or T.

36 A panel of primers comprising 10 primers of the sequences A-G-G-T-C-G-A-C-G-G-T-A-T-C-G-G-V-N-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide

-56-

selected from the group consisting of A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T.

5 37. A degenerate mixture of primers comprising a mixture of 12 primers of the sequences A-A-C-T-G-G-A-A-G-A-A-T-T-C-G-C-G-G-C-C-G-C-A-G-G-A-A-T-T-T-T-T-T-T-T-T-T-T-V-N (SEQ ID NO: 2), wherein V is a deoxyribonucleotide selected from the group consisting of
10 A, C, and G; and N is a deoxyribonucleotide selected from the group consisting of A, C, G, and T, each of the 12 primers being present in about an equimolar quantity.

1/2

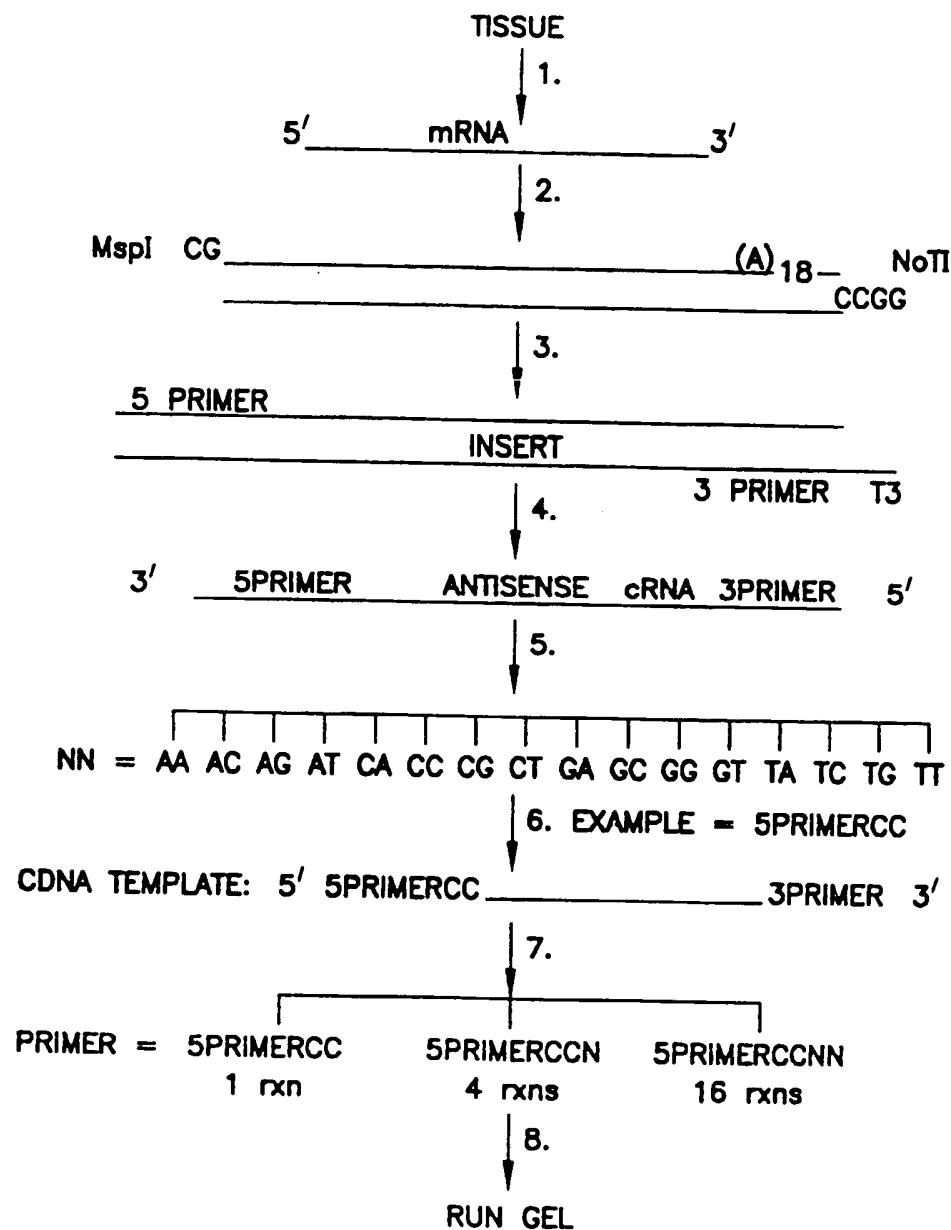


FIG 1

SUBSTITUTE SHEET (RULE 26)

2/2

106→130 CX } NSE
 116→134 LV }
 118→111 CX } RC3
 118→127 LV }
 118→128 CX } SOMAT.
 LV }

500

400

344

300

220

200

RT'sPCR's

154

128 - CTAC
 118 - CT → 127 - CTGA
 111 - CTGC

134

116 - GT → 134 - GTGC
 106 - CG → 130 - CGGC

2 2 2

FIG. 2

SUBSTITUTE SHEET (page 20)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/13041

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 15/11; C12P 19/34; C12Q 1/68
US CL :435/6, 91.1, 91.3; 536/24.33

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1, 91.3; 536/24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog
search terms: PCR, mRNA, display, pattern

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STRATAGENE CLONING SYSTEMS, Product Catalog, published 1993, page 312, see entire product description.	33-35
Y	Science, Volume 257, issued 14 August 1992, Liang et al, "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction", pages 967-971, see entire article.	36, 37
Y	Nucleic Acids Research, Volume 21, Number 4, issued 1993, Liang et al, "Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization", pages 3269-3275, see entire article.	36, 37

<input checked="" type="checkbox"/>	Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.
* A	Special categories of cited documents:
"A"	document defining the general state of the art which is not considered to be of particular relevance
"E"	earlier document published on or after the international filing date
"L"	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O"	document referring to an oral disclosure, use, exhibition or other means
"P"	document published prior to the international filing date but later than the priority date claimed
"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&"	document member of the same patent family

Date of the actual completion of the international search | Date of mailing of the international search report

14 APR 1995

14 FEB 1995

Name and mailing address of the ISA
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Faximile No. (703) 305-3230

SEARCHED

PHILIP W. CARTER

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/US94/13041**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Trends in Genetics, Volume 5, Number 6, issued June 1989, White et al, "The polymerase chain reaction", pages 185-189, see entire article.	36, 37